Problem formulation and phenotypic characterisation for the development of novel crops

NB Change volume number for Sanvido et al. 2012 from 9 to 15

Problem formulation and phenotypic characterisation for the development of novel crops

Alan Raybould, Syngenta Crop Protection AG, Schwarzwaldallee 215, 4058 Basel, Switzerland

alan.raybould@syngenta.com

Abstract

Phenotypic characterisation provides important information about novel crops that helps their developers to make technical and commercial decisions. Phenotypic characterisation comprises two activities. Product characterisation checks that the novel crop has the qualities of a viable product – the intended traits have been introduced and work as expected, and no unintended changes have been made that will adversely affect the performance of the final product. Risk assessment evaluates whether the intended and unintended changes are likely to harm human health or the environment. Product characterisation follows the principles of problem formulation, namely that the characteristics required in the final product are defined and criteria to decide whether the novel crop will have these properties are set. The hypothesis that the novel crop meets the criteria are tested during product development. If the hypothesis is corroborated, development continues, and if the hypothesis is falsified, the product is redesigned or its development is halted. Risk assessment should follow the same principles. Criteria that indicate the crop poses unacceptable risk should be set, and the hypothesis that the crop does not possess those properties should be tested. However, risk assessment, particularly when considering unintended changes introduced by new plant breeding methods such as gene editing, often ignores these principles. Instead, phenotypic characterisation seeks to catalogue all unintended changes by profiling methods and then proceeds to work out whether any of the changes are important. This paper argues that profiling is an inefficient and ineffective method of phenotypic characterisation for risk assessment. It discusses reasons why profiling is favoured and corrects some misconceptions about problem formulation.

Key words: hypothesis testing; decision-making; product characterisation; risk assessment; plant breeding; profiling

Introduction

The aim of crop breeding is to introduce useful heritable traits into crops while minimising adverse unintended effects. Breeding methods vary enormously, and include making hybrids between different species, treating crops with radiation or chemicals to induce numerous untargeted mutations, using recombinant DNA techniques to move genes between species, and gene editing, which can make targeted changes to a single nucleotide (Pacher and Puchta 2017). Whatever the breeding method used in their production, new crops designed to be commercial products require phenotypic characterisation to check that the intended changes have been made and that they have the desired effects. In addition, the crops must be assessed for the presence of unintended changes that would make the product inviable.

Phenotypic characterisation of a novel crop is comparative. The new crop may be compared with an existing variety, as is routine in the evaluation of genetically modified (GM) crops (Nickson 2008). The crop may also be compared against a standard; for example, to be classed as Canola, rapeseed varieties must meet certain requirements about glucosinolate content (Sang and Salisbury 1988). In addition, the awarding of Plant Breeders' Rights for a new variety depends on its being distinct from all previously released varieties, as well as being uniform and stable (Cockram et al. 2012).

Phenotypic characterisation of novel crops helps decision-making during product development. It informs a succession of checks that the product meets certain design specifications. If the product meets the specifications, development is continued; if it fails to meet them, development is discontinued or the product is redesigned by further breeding.

This paper argues that decision-making in crop development is best served by problem formulation that sets decision-making criteria first and then uses phenotypic characterisation to test the hypothesis that the potential product meets these criteria. Decision-making is not well served by making a detailed description (profile) of the crop and any comparator, and then trying to evaluate the relevance of differences in the profiles.

The following discussion of the merits of problem formulation and profiling concentrates on decision-making by product developers. One reason for this is to separate discussion of problem formulation from regulation of new crops. All new crops require some degree of phenotypic characterisation to help their developers make sound technical and commercial decisions. Hence, the suggestion that problem formulation is useful for phenotypic characterisation of a crop is not a suggestion that the crop ought to be subject to pre-market regulation. This clarification is particularly important for crops produced by gene editing because their regulatory status is controversial and depends on legal definitions, not on expectations about their phenotypes (Kupferschmidt 2018).

Profiling is an empirical or data-driven approach to science that tries to eliminate bias (Kok and Kuiper 2003), whereas problem formulation is hypothesis-driven (Raybould 2006). We first examine problems with empiricism and how hypothesis-driven science offers solutions. Next we use a simple imaginary example to highlight important differences between profiling and hypothesis testing to support decision-making. We then examine how these ideas may be applied to phenotypic characterisation of crops.

Data-driven and hypothesis-driven science

Differences in philosophy

When modern science emerged in late Sixteenth Century, a common view was that objective knowledge arises from observations made without preconceptions. This approach, called empiricism, was supposed to guard against subjectivity and prejudice and thereby develop

an alternative to religious interpretations of nature (e.g., Hahn 1965, Ayala 2009). Patterns in the observations lead to hypotheses, which are proved to be true knowledge when sufficient confirmatory observations have accumulated (e.g., Phelan 2001).

Popper (1989) argued that observing without preconceptions is impossible: “Observation is always selective. It needs a chosen object, a definite task, an interest, a point of view, a problem.” In addition, philosophers had long worried about the logical impossibility of proving the truth of hypotheses by observation because future observation may show them to be wrong (e.g., Lantin 1998). Popper (1979) developed his insight about observations to argue that objective knowledge grows by our attempts to solve problems. In science, we propose hypotheses as solutions to problems raised by unexpected observations, and test them by examining the accuracy of their predictions. If the predictions of the hypothesis agree with our observations, the hypothesis is corroborated. If predictions and observations disagree, the hypothesis is falsified. Knowledge is provisional and grows by correcting errors in the light of observations.

Popper’s concept of knowledge production led him to devote much effort to describing the properties of good hypotheses: they should seek not only accuracy but also high informative content. A hypothesis that predicts rain somewhere in the world next month is likely to be accurate, but has low informative content – it excludes very few observations. A hypothesis that there will be 2 cm of rain in London next Friday is less likely to be accurate, but has higher informative content. Popper (1979) developed a concept called verisimilitude, combining accuracy and information content, to judge the quality of a hypothesis. While Popper’s formal theory of verisimilitude has been criticised (Thornton 2018), his argument that good hypotheses have high informative content and make precise, and hence testable and improbable, predictions is generally accepted (Chalmers 2013). On these grounds, a hypothesis predicting the amount of rain in London next Friday is better than the hypothesis predicting some rain somewhere in the world next month.

Differences in application

Popper was writing about basic science and so his interest was devising hypotheses that are useful for producing knowledge; in philosophical terms, they have high epistemic utility. Applied science, on the other hand, seeks to develop tools that help us to achieve our goals (Niiniluoto 1994). Hence, although epistemic utility is important in applied science, simplicity and manageability that provide high practical utility are also crucial criteria for judging hypotheses (Niiniluoto 1993).

In essence, problem formulation is concerned with maximising practical utility for decision-making and does so by devising hypotheses that specific criteria are met. If we must decide whether to pack an umbrella for a visit to London, we may want to know the probability of rain in Hyde Park between 2pm and 4pm tomorrow. Suppose our existing knowledge indicates that falling barometric pressure at Heathrow Airport means the probability of rain tomorrow afternoon in Hyde Park is greater than 50%, while steady or rising pressure means the probability of rain is less than 50%. A hypothesis with high practical, but low epistemic, utility is that barometric pressure at Heathrow at 2 pm today is greater than or equal to the pressure at Heathrow at 2 pm yesterday. If the hypothesis is corroborated, we leave the umbrella at home; if it is false, we pack the umbrella.

Profiling is more concerned with detailed descriptions than hypotheses; indeed, many scientists consider profiling to be a hypothesis-free, data-driven approach to science (Vlaanderen et al. 2010). In the umbrella example, profiling might build a picture of the differences between today’s and yesterday’s weather at Heathrow – perhaps comparing second-by-second variations in temperature, wind speed, cloud cover, humidity, barometric

pressure and precipitation. However, in choosing to compare weather data, profiling at least has a hypothesis that these data are more useful than, say, data on traffic, illustrating Popper's point that one cannot observe without a point of view.

Profiling usually goes on to test a null hypothesis of no difference between the things being compared. However, such hypotheses are means of presenting data rather than hypotheses derived from theory (Stephens et al. 2007); hence, they have low epistemic utility. Null hypotheses also have low practical utility. If we have no decision-making criteria – such as a 50% probability of rain in Hyde Park tomorrow – we have no way to determine which, if any, of the inevitably numerous differences between today's and yesterday's weather are relevant to our decision whether to pack an umbrella.

If profiling moves from cataloguing differences to explaining differences – such as the physical mechanisms by which today's weather leads to tomorrow's – it starts creating hypotheses that have high epistemic utility. However, without decision-making criteria, we cannot determine which have practical utility. The mistake of profiling is to think that decision-making criteria can be derived from detailed descriptions or explanations. The umbrella example shows this to be false. Using 50% probability of rain as our decision-making criterion for taking an umbrella is determined by the values we place on staying dry and on not carrying an umbrella unnecessarily. Someone who values being dry more than we do may set a much lower probability of rain as the trigger for taking an umbrella. These values cannot be derived from descriptions or explanations of the weather, however detailed and accurate they may be. Similarly, detailed phenotypic profiles of a novel crop and a comparator, or a complicated model explaining the reasons for differences in the profiles, cannot determine which differences are important. We can determine which differences are important only by first deciding what we want to achieve and the properties of a novel crop that contribute to achieving those objectives.

Application of problem formulation to phenotypic characterisation

Phenotypic characterisation for decision-making in the development of a novel crop is analogous to the umbrella problem, above. There is a clear problem, namely whether to continue development of the product. Then there are the components of the developer's decision, which may include the probability that using the product will provide the intended benefits for customers and hence be profitable (the opportunity to the developer from developing the product), the probability that using the product will cause specified harms and hence create liabilities or reputational damage (the risk to the developer from developing the product), and how the developer should weigh opportunity and risk (Sanvido et al. 2012). If the developer decides that the potential product provides acceptable opportunity if it shows at least a 5% improvement in insect control over the comparator under identical conditions, then the hypothesis that there is at least a 5% improvement in insect control should be tested. The final part of problem formulation is producing a plan to test the hypothesis.

A vital conclusion from equivalence of the umbrella and product development problems is that phenotypic characterisation for product development should not indulge in profiling or explanation. There are many techniques for profiling novel crops and comparators, including measuring morphology and development (Horak et al. 2015, Tardieu et al. 2017), compositional analysis (Rayan and Abbott 2015), various omics methods (Davies 2010, Ricroch et al. 2011, Li et al. 2017) and mass spectrometry (García-Cañas et al. 2011). At some level of detail, a novel crop under evaluation will differ from its comparator in ways that are intended and unintended by the breeder. However, product development will not be helped by exhaustively cataloguing or explaining these differences, only by testing for differences that affect a decision.

We now look at how problem formulation may be applied to different product development decisions. As stated above, to avoid implying that products of a particular technology should or should not be regulated, we will discuss decision-making from the point of view of the crop developer, not a regulator. However, the same principles apply to regulatory decision-making. We also separate product characterisation, which determines whether the product works as intended, from risk assessment, which determines whether use of the product will harm human health or the environment. Conceptually, there is overlap between product characterisation and risk assessment: if a product poses significant risk to, say, human health, it is probably not working as intended. However, we consider these topics separately to reflect common practice in crop development.

Product characterisation

Product characterisation usually deals with attributes of the crop that are useful to users of the product. These may be agronomic characteristics desired by farmers or quality characteristics valued by consumers (Schaart et al. 2016, Francis et al. 2017). Product characterisation has two elements: a test of the hypothesis that the intended trait has been introduced and performs as intended; and a test of the hypothesis that other agronomically and nutritionally important characters have not been inadvertently changed in a way that makes the product inviable. If the crop is regulated, product characterisation will also check that certain regulatory standards are met or that certain characters that may raise regulatory concerns are not present; for example, GM crops are checked for the absence of multiple copies of the transgene cassette (Prado et al. 2014). Unintentional changes that may be harmful to human health and the environment are usually considered separately in risk assessments.

Product characterisation uses problem formulation. The crop developers will have a clear target, perhaps to produce a new variety with increased drought tolerance, and well-defined criteria for predicting success of the final product, such as a certain increase in yield over a competitor variety in dry conditions and no decrease in yield under optimum conditions (Tester and Langridge 2010). The suitability of the finished new variety will be determined by testing the hypothesis that it meets these criteria when used commercially.

Depending on the breeding technique used, there may be a series of intermediate decisions each with its own criteria, hypotheses and tests leading up to a test of the finished variety. If the improved drought tolerance is introduced by genetic modification, the initial transformants may be screened in the laboratory to test the hypothesis that they have integrated the transgene and that it confers some basic improvement in water-use efficiency (Deikman et al. 2012). More realistic hypotheses will be tested to whittle down the potentially many transformed plants to one or a few candidate products (Deikman et al. 2012, Prado et al. 2014). Similarly, if improved drought tolerance were being sought by conventional breeding, breeders may test the hypothesis that seedlings contain genetic markers linked to genes that confer high water-use efficiency in mature plants (Cattivelli et al. 2008).

As well as intended effects, breeding introduces unintended effects that may arise through pleiotropic or epistatic effects of the genetic changes underlying production of the intended trait (Miedaner and Korzun 2012). Unintended effects may also arise as a result of the breeding method used; for example, linkage drag with conventional breeding (Zamir 2001), numerous random mutations with mutation breeding (Brunner 1995), insertional mutagenesis with genetic modification (Cellini et al. 2004), and off-target edits with gene editing (Svitashev et al. 2016). Product characterisation tends not to use profiling for evaluating unintended effects because not all unintended effects are necessarily adverse to the viability of the product. Instead, product characterisation tests that agronomic and nutritional properties of the crop known to be important to farmers or consumers are within acceptable limits.

In general, agronomic assessments measure germination, establishment, growth and development, including flowering time, pest and disease resistance and yield. The characters measured and their relative importance in decision-making vary among crops; further details of agronomic assessments of maize, rice, soybean and what are given by Hallauer et al. (1988), Khush (1987), Morrison et al. (2000) and Cox et al. (1987), respectively. Nutritional assessments may include the content of minerals, vitamins and fatty acids, and any anti-nutrients or toxins known to be produced by the crop species in question. Further details and an introduction to the literature are given by Ridley et al. (2004) and Alba et al. (2010).

The purpose of the product characterisation is to assess risks to the developer if the crop were commercialised. Although data from agronomic and nutritional analysis trials may be used in human health and environmental risk assessment (see below), the failure of a crop to perform well in these studies (e.g., Zeller et al. 2010) should not be assumed to indicate that it poses health or environmental risks – many of the characters measured have no relevance to these risks. In the case of ecological risks posed by the weediness or invasiveness potential of the crop, failure of the crop to grow well usually means that the risk is lower than that posed by better-performing varieties (Ellstrand et al. 2010).

Finally, while it is not necessary to explain all the unintended effects revealed in product characterisation studies, knowledge of the genetic or molecular basis of adverse effects that would prevent further development of the product may be useful if they provide a strategy for mitigation. Knowledge of the inheritance of unwanted traits, for example, might help to determine whether they could be removed by backcrossing (Xu and Crouch 2008). The usefulness of studies to investigate the mechanism underlying unwanted traits does not invalidate previous arguments about profiling and explanation in product characterisation. Studies of mechanism are conducted in response to a decision that the product in its present form is unsatisfactory, not to set criteria for making that decision.

Risk assessment

Risk assessment is similar to product characterisation, but its focus is identifying attributes of the crop that may lead to harm to human health or the environment. Definitions of harm may be derived from relevant laws and any regulations specific to the type of product being assessed (Sanvido et al. 2012, Garcia-Alonso and Raybould 2014). Product developers would run severe risks to their institutions by commercialising products that are likely to cause such harm. In the case of products requiring pre-market authorisation, they would probably not receive approval to sell the product. In addition to legal definitions, developers may extend the definitions of harm as a form of self-regulation in compliance with their business ethics (Gunningham and Rees 1997).

Some aspects of phenotypic characterisation for risk assessment are driven by problem formulation and hence are similar to product characterisation. If the intended trait in the plant is, say, the production of a new protein that is toxic to insect pests, phenotypic characterisation would measure the concentration of the protein in various plant tissues. These data may be combined with ecotoxicology and environmental fate data to determine various ecological risks (Garcia-Alonso et al. 2006, Romeis et al. 2008). Decision-making criteria are usually the ratios of effects measured in ecotoxicology studies, such as the no observed adverse effect concentrations for a variety of organisms, and the predicted exposures of those types of organism in the field. These exposures are based on the measured concentrations in the crop tissues, along with information about the route of exposure of the organism concerned (Head et al. 2001). The value of the ratio that indicates acceptable risk is usually set by regulations, not the product developer (US EPA 2007). Conceivably, a product developer could set its own criteria based solely on the amount of

protein in particular tissues, perhaps in response to concerns about unwanted exposure regardless of the likely effect of that exposure (Wunderlich and Gotto 2015).

Another aspect of problem formulation for phenotypic characterisation is assessment of defined components of crops for food safety. If breeders have made intended compositional changes, such as increasing the concentration of lycopene in tomatoes, an assessment of the potential for harmful side effects of that change may be made (Kok et al. 2008). Also, if the crop has a history of producing substances harmful to health, checking that breeding has not unintentionally raised the concentration of those substances above a threshold of concern is a sensible precaution; such substances include furanocoumarins in celery (Berkley et al. 1986), glycoalkaloids in potatoes (Camire et al. 2009) and glucosinolates in *Brassica* vegetables and oilseed crops (Jahangir et al. 2009).

Other elements of phenotypic characterisation for risk assessment, particularly those requested by regulatory authorities, deal with unintended effects, and are driven by profiling rather than by problem formulation. Applicants for premarket approval of a GM crop must submit data comparing the crop's composition and agronomic performance with those of a suitable comparator. While these data may be useful in product characterisation, because the measured characters can usually be directly related to product performance and hence to decision-making criteria, there are few, if any, similar criteria for judging whether changes in these characters pose unacceptable risks to human health or the environment. Hence in risk assessment, agronomic and compositional analyses are simply profiling – they describe the GM crop rather than test hypotheses about the acceptability of the risks its use poses.

Profiling leads to problems. A large, well-executed agronomic or compositional analysis study that measures many variables will almost certainly detect statistically significant differences between the crop being assessed and its comparator. Much time and effort may be spent attempting to work out the biological relevance of these differences. However, in the context of risk assessment, the biological relevance of a difference should be determined by whether it falsifies a hypothesis that use of the crop poses acceptable risk (a hypothesis with high practical utility), not whether it is interesting scientifically (leading to the development and testing of hypotheses with high epistemic utility) (Hill and Sendashonga 2003). If a decision-maker cannot state the size of difference in a character, or combination of characters, that would give concern, then measuring those characters has no value for risk assessment. Indeed, the effort to collect and analyse the data may distract from the design and interpretation of experiments that would be useful for decision-making (Raybould 2010). The introduction of omics techniques for profiling in risk assessment (e.g., Davies 2010) would make matters worse (Raybould and Macdonald 2018).

Problem formulation provides a different approach. First, unacceptable effects of using the crop would be defined; this may be more straightforward for human health, where increases in morbidity and mortality are almost universally accepted as harmful, than for environmental effects (Sanvido et al. 2012, Devos et al. 2014). Next, differences between a novel crop and its comparator that indicate an unacceptable probability of these effects occurring would be defined. Then the hypothesis that these differences do not occur would be tested. If the hypothesis is falsified, further analysis of the differences may be conducted, or development of the crop could be stopped.

Defining unacceptable differences and testing the hypothesis that they do not occur is criticised as a biased (Kok et al. 2008), naïve (Arpaia et al. 2017) or even hubristic (Wynne 2001) approach to risk assessment (see below). In consequence, profiling is often recommended as a better option for evaluating unintended effects in novel crops. However, using profiling for product characterisation would appear to be irrational. No product developer would profile a new crop in the hope that unintended opportunities might be

revealed; instead, they would define product performance criteria and test the hypothesis that the potential new product meets those criteria. Problem formulation for risk assessment is only different in that it defines attributes of the crop that should not be present rather than those that should be present. In the next section, we discuss why profiling unintended effects may be suggested for risk assessment but not for product characterisation.

Risk assessment, profiling and unintended effects

A recurrent theme in risk assessment of novel crops is the need to assess unintended effects caused by the breeding method, particularly if the breeding method is new. All breeding methods have the potential to introduce unintended effects; hence, product characterisation for all new crops will check that no specific unwanted changes have been unintentionally introduced (see above). In risk assessment, however, changes that would cause concern tend not to be specified. Instead, attempts are made to catalogue unintended changes using profiling. This method is particularly recommended for newer methods of plant breeding, such as genetic modification and gene editing (Kuiper et al. 2003, Wolt 2017). Below we discuss four reasons for the support for profiling. These reasons are not necessarily comprehensive or independent.

Establishing similarity

A simple reason for using profiling in risk assessment may be an idea that establishing similarities between a new crop and a comparator demonstrates low risk, and the more similarities that are demonstrated, the lower the risk. This idea is related to the concept of familiarity (Madsen et al. 2002), which draws on the reasonable idea that knowledge of existing crops can be used in the risk assessment of novel crops. This does not mean, however, that risk assessment has to establish similarity between the new crop and its comparator for as many characters as can be measured, or that every non-statistically significant difference is further evidence for low risk. Low risk is established by testing and corroborating hypotheses that there are no *potentially harmful* differences between the crop and the comparator, not by some measure of similarity of their profiles.

A simple example demonstrates the idea that the number of of similarities cannot demonstrate low risk. A novel crop having 10 times the concentration of an endogenous toxin of a comparator, but showing no statistically significant differences in 99 measures of agronomic performance, would pose greater risk to health than would a novel a crop that has the same the toxin concentration as the comparator, but differs from it on all 99 measures of agronomic performance. Existing knowledge is useful for defining what differences are potentially harmful, such as increases in the concentration of toxin. It should not be a baseline for exhaustively testing null hypotheses of no unintended changes in the novel crop.

A focus on establishing similarity between a novel crop and its comparator can lead to unproductive debates and assessments. Expanded versions of agronomic and crop quality assessments done for product characterisation are now routinely considered as part of regulatory risk assessment even though many of the characters measured in those studies are unrelated to processes that could cause harm to health or the environment. In effect, studies designed by problem formulation for product characterisation become profiling studies when used for risk assessment.

Another problem with similarity driving risk assessment is that considerations of the statistics of comparative studies, such whether to test for difference or equivalence (Perry et al. 2009, van de Voet 2011), can crowd out discussion of what differences are potentially harmful. Discussion of the most suitable statistical test for analysing certain data is only relevant for risk assessment if the data themselves are relevant for risk assessment.

Avoiding bias

A second reason for advocating profiling is the idea that risk assessment should be unbiased (Kok et al. 2008). Such thinking is a throwback to the foundation of empiricism (see above), when preconceptions were considered inimical to sound observations. Part of the problem is, perhaps, the term "bias", which implies improper behaviour. Popper's more neutral term, "having a point of view", might be more suitable. Whatever term is used, risk assessment cannot function without a point of view. Harm must be defined based on the decision-maker's values and objectives, and from such definitions, decision-making criteria can be defined and tests that the criteria are met can be performed. Eliminating such "bias" leads to risk assessment operating in a vacuum with no way to judge which data are valuable (Evans et al. 2006). Profiling is the almost inevitable result.

Difficulty in defining indicators of harm

A third reason for recommending profiling may be that defining what differences between a novel crop and its comparator are potentially harmful is thought to be too difficult. This reason is inadequate for at least two reasons. First, there is a wealth of knowledge on how crops may harm human health (Ory 1981), or become weeds of agriculture and invasive plants outside agriculture (Gressel 2004). Exploring this literature to identify crop characters most likely to led to harm can be fruitful for problem formulation (e.g., Novak and Haslberger 2000, Devos et al. 2018), especially if the crop breeding method is thought not to introduce new hazards (NASEM 2016). Second, even if existing knowledge is thought inadequate for identifying potentially harmful effects and decision-making criteria, it is not clear how profiling will improve matters. If there is reason to believe that the new crop may contain a toxin previously unknown in that species, or trait that increases its ecological fitness, then that should lead to a hypothesis about its occurrence that can be tested by targeted experiments.

Unknown unknowns and unintended consequences

A final reason for choosing profiling may be a response to criticism that hypothesis-led risk assessment ignores "unknown unknowns" and is hubristic (Wynne 2001). A related view is that predicting the consequences of using novel crops is inevitably highly complex and uncertain and therefore the only sensible approach is a precautionary ban on using novel crops until the consequences of their use are "fully characterised" (Myhr 2010). By measuring many things without ascribing values to them, profiling may be seen as more capable of fully characterising consequences, and therefore more humble and more open to complexity and unexpected consequences.

Every activity will be associated with unknown unknowns and by definition they are not amenable to scientific analysis. Unknown unknowns, therefore, cannot justify scientific arguments to favour one activity (e.g., not using a novel crop) over another (e.g., using the crop), although they may underlie psychological or political reasons for preferring not to use products of new agricultural technologies (Herring and Paarlberg 2016). If the unknown unknowns of one activity are thought to be worse than another, then we must have some knowledge of their likely consequences, meaning that the unknowns are not unknown.

Also, every activity will have unintended consequences and attempting to predict them all is inevitably complex. Using an umbrella in London may have enormous consequences; for example, someone might miss a vital meeting if they miss a bus because its number is obscured by our umbrella. Attempting to envisage all the consequences of just that event – and there maybe thousands of similar events – is not only extremely difficult, but probably also worthless. Even if one had perfect knowledge of the consequences of one's using and not using an umbrella in London, how could values be ascribed to them all so that this knowledge could be used in decision-making? The same reasoning applies to unintended

differences between a novel crop and a comparator – comprehensively cataloguing those differences, working out their consequences if the product were used or not used, and ascribing values to those consequences would prevent effective decision-making.

Problem formulation does not claim to lead to scientific certainty or "full characterisation" of a proposed action. Instead, it sets out decision-making criteria based on the policy objectives of whoever is making the decision and devises tests of the hypothesis that the criteria are met. These criteria cannot and should not hope to incorporate all consequences of an action, only those that are most important to a decision. There may be legitimate disagreements over who is entitled to a say in setting the decision-making criteria (Jasanoff and Hurlbut 2018); however, that does not mean that risk assessment should proceed without decision-making criteria and use untargeted profiling instead.

Uses of profiling in technology evaluation

Rejecting profiling for risk assessment does not imply that it is not useful in other activities. Comparing the incidence of unintended changes associated with different breeding methods may be useful for evaluating their relative precision and accuracy (Pacher and Puchta 2017). This knowledge will help breeders when choosing a method to produce a desired crop phenotype, provided they have clear decision-making criteria based on the incidence of the type of change measured. Profiling may also be valuable for testing hypotheses about how to reduce the frequency of unintended changes when using a particular breeding method (Yin et al. 2017). Again, the practical application of such an approach will depend on having a clear criterion for the acceptability of the method that is based on the frequency of unintended changes of the type measured by the profiling. In both cases, profiling is used to test hypotheses developed by problem formulation, not as a substitute for problem formulation.

Conclusion: in favour of hypothesis-driven phenotypic characterisation

Problem formulation sets decision-making criteria and devises tests of hypotheses that those criteria are met. Its use in phenotypic characterisation of novel crops for product characterisation is common. Its use in phenotypic characterisation for risk assessment is much rarer and is under pressure from political and scientific trends. A fashion for evidence-based policymaking as an impartial way to decide "what works" (Clarence 2002) means that problem formulation may be seen as unacceptably biased. In addition, big data is encouraging the view that collecting data is an alternative to testing hypothesis as the route to knowledge (Kell and Oliver 2004).

Problem formulation is a tool for helping to determine what works, but it cannot be done effectively by avoiding definition of what "working" entails. In the context of the development of novel crops, effective problem formulation defines what using the crop is intended to achieve, what side-effects of using the crop should be avoided, and thereby defines criteria that must be met for product development to continue. The decisions themselves must rest on the results of testing hypotheses that the product meets those criteria. These concepts work well in phenotypic characterisation for product characterisation and are uncontroversial. Their use in phenotypic characterisation for risk assessment should also work well because it is essentially the same process as product characterisation, but substituting criteria for avoiding unacceptable risk to human health and the environment for criteria that predict whether the product will be commercially successful. The setting of decision-making criteria in risk assessment is not improper bias, and big data does not make hypothesis testing an obsolete or hubristic method of producing the knowledge necessary for decision-making. Problem formulation for risk assessment should be welcomed as an opportunity to focus phenotypic characterisation on characters that really matter for decision-making. It should

not be rejected in favour of profiling methods that deliberately shy away from making judgements about whether certain data are useful.

References

Alba R, Phillips A, Mackie S, Gillikin N, Maxwell C, Brune P, Ridley W, Fitzpatrick J, Levine M, Harris S (2010) Improvements to the International Life Sciences Institute Crop Composition Database. J Food Compos Anal 23: 741–748

Arpaia S, Birch ANE, Kiss J, van Loon JJA, Messéan A, Nuti M, Perry JN, Sweet JB, Tebbe CC (2017) Assessing environmental impacts of genetically modified plants on non-target organisms: the relevance of in planta studies. Sci Total Environ 583: 123–132

Ayala FJ (2009) Darwin and the scientific method. Proc Natl Acad Sci USA 106: 10033–10039

Berkley SF, Hightower AW, Beier RC, Fleming DW, Brokopp CD, Ivie GW, Broome CV (1986) Dermatitis in grocery workers associated with high natural concentrations of furanocoumarins in celery. Ann Intern Med 105: 351–355

Brunner H (1995) Radiation induced mutations for plant selection. Appl Radiat Isot 46: 589–594

Camire ME, Kubow S, Donnelly DJ (2009) Potatoes and human health Crit Rev Food Sci Nutr 49: 823–840

Cattivelli L, Rizza F, Badeck F-W, Mazzucotelli E, Mastrangelo AM, Francia E, Marè C, Tondelli A, Stanca AM (2008) Drought tolerance improvement in crop plants: an integrated view from breeding to genomics. Field Crop Res 105: 1–14

Cellini F, Chesson A, Colquhoun I, Constable A, Davies HV, Engel KH, Gatehouse AMR, Kärenlampi S, Kok EJ, Leguay J-J, Lehesranta S, Noteborn HPJM, Pedersen J, Smith M (2004) Unintended effects and their detection in genetically modified crops. Food Chem Toxicol 42: 1089–1125

Chalmers AF (2013) What is this this thing called science? 4th edition. Hackett Publishing, Indianapolis

Clarence E (2002) Technocracy reinvented: the new evidence based policy movement. Public Policy Adm 17: 1–11

Cockram J, Jones H, Norris C, O'Sullivan DM (2012) Evaluation of diagnostic molecular markers for DUS phenotypic assessment in the cereal crop, barley (*Hordeum vulgare* ssp. *vulgare* L.). Theor Appl Genet, 125: 1735–1749

Cox TS, Shroyer JP, Ben-Hui L, Sears RG, Martin TJ. (1987) Genetic Improvement in agronomic traits of hard red winter wheat cultivars 1919 to 1987. Crop Sci 28: 756–760

Davies H (2010) A role for "omics" technologies in food safety assessment. Food Control 21: 1601–1610

Deikman J, Petracek M, Heard JE (2012) Drought tolerance through biotechnology: improving translation from the laboratory to farmers' fields. Curr Opin Biotechnol 23: 243–250

Devos Y, Ortiz-García S, Hokanson KE, Raybould A (2018) Teosinte and maize × teosinte hybrid plants in Europe – environmental risk assessment and management implications for genetically modified maize. Agric Ecosyst Environ 259: 19–27

Devos Y, Sanvido O, Tait J, Raybould A (2014) Towards a more open debate about values in decision-making on agricultural biotechnology. Transgenic Res 23: 933–943

Ellstrand NC, Heredia SM, Leak-Garcia JA, Heraty JM, Burger JC, Yao L, Nohzadeh-Malakshah S, Ridley CE (2010) Crops gone wild: evolution of weeds and invasives from domesticated ancestors Evol Appl 3: 494–504

Evans J, Wood G, Miller A (2006) The risk assessment – policy gap: an example from the UK contaminated land regime. Environ Int 32: 1066–1071. https://doi.org/10.1016/j.envint.2006.06.002

Francis D, Finer JJ, Grotewold E (2017) Challenges and opportunities for improving food quality and nutrition through plant biotechnology. Curr Opin Biotechnol 44: 124–129

Garcia-Alonso M, Jacobs E, Raybould A, Nickson TE, Sowig P, Willekens H, Van der Kouwe P, Layton R, Amijee F, Fuentes A, Tencalla F (2006) Assessing the risk of genetically modified plants to non-target organisms. Environ Biosafety Res 5: 57–65

Garcia-Alonso M, Raybould A (2014) Protection goals in environmental risk assessment: a practical approach. Transgenic Res 23: 945–956

García-Cañas V, Simó C, León C, Ibáñez E, Cifuentes A (2011) MS-based analytical methodologies to characterize genetically modified crops. Mass Spectrom Rev 30: 396–416

Gressel J (ed.) (2004) Crop ferality and volunteerism. CRC Press, Boca Raton

Gunningham N, Rees J (1997) Industry self-regulation: an institutional perspective. Law Policy 19: 363–414

Hahn R (1965) Reflections on the history of science. J Hist Philos 3: 235–242

Hallauer AR, Russell WA, Lamkey KR (1988) Corn breeding. In: Sprague GF, Dudley JW (eds) Corn and corn improvement. Third edition. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, pp 463–564

Head G, Brown CR, Groth ME, Duan JJ (2001) Cry1Ab protein levels in phytophagous insects feeding on transgenic corn: implications for secondary exposure risk assessment. Entomol Exp Appl 99: 37–45

Herring R, Paarlberg R (2016) The political economy of biotechnology. Ann Rev Resour Econ 8: 397–416

Hill RA, Sendashonga C (2003) General principles for risk assessment of living modified organisms: lessons from chemical risk assessment. Environ Biosafety Res 2: 81–88

Horak MJ, Rosenbaum EW, Kendrick DL, Sammons B, Phillips SL, Nickson TE, Dobert RC, Perez T (2015) Plant characterisation of Roundup Ready 2 Yield® soybean, MON 89788, for use in ecological risk assessment. Transgenic Res 24: 213–225

Jahangir M, Kim HK, Choi YH, Verpoorte R (2009) Health-affecting compounds in *Brassicaceae*. Comp rev Food Sci Food Saf 8: 31–43

Jasanoff S, Hurlbut JB (2018) A global observatory for gene editing. Nature 555: 435–437

Kell DB, Oliver SG (2004) Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. BioEssays 26: 99–105

Khush GS (1987) Rice breeding: past and present. J Genet 66: 195–216

Kok EJ, Keijer J, Kleter GA, Kuiper HA (2008) Comparative safety assessment of plant-derived foods. Regul Toxicol Pharmacol 50: 98–113

Kok EJ, Kuiper HA (2003) Comparative safety assessment for biotech crops. Trends Biotech 21: 439–444

Kuiper HA, Kok EJ, Engel K-H (2003) Exploitation of molecular profiling techniques for GM food safety assessment. Curr Opin Biotech 14: 238–243

Kupferschmidt K (2018) EU verdict on CRISP crops dismays scientists. Science 361: 435–436Li

Lantin R (1998) Hume and the problem of induction. Philosophia 26: 105–107

Li R, Quan S, Yan X, Biswas S, Zhang D, Shi J. Molecular characterization of genetically-modified crops: challenges and strategies. Biotechnol Adv. 2017; 35: 302–309.

Madsen KH, Holm PB, Lassen J, Sandøe P (2002) Ranking genetically modified plants according to familiarity. J Agric Environ Ethics 15: 267–278

Miedaner T, Korzun V (2012) Marker-assisted selection for disease resistance in wheat and barley breeding. Phytopathology 102: 560–566

Morrison MJ, Voldeng HD, Cober ER (2000) Agronomic changes from 58 years of genetic improvement of short-season soybeans in Canada. Agron J 92: 780–784

Myhr AI (2010) A precautionary approach to genetically modified organisms: challenges and implications for policy and science. J Agric Environ Ethics 23: 501–525

NASEM (National Academies of Sciences, Engineering, and Medicine) (2016) Genetically engineered crops: experiences and prospects. National Academies Press, Washington, DC

Novak WK, Haslberger AG (2000) Substantial equivalence of antinutrients and inherent plant toxins in genetically modified novel foods. Food Chem Toxicol 38: 473–483

Nickson TE (2008) Planning environmental risk assessment for genetically modified crops: problem formulation for stress tolerant crops. Plant Physiol 147: 494–502

Niiniluoto I (1993) The aim and structure of applied research. Erkenntnis 38: 1–21

Niiniluoto I (1994) Approximation in applied science. In: Kuokkanen M (ed) Idealization VII: structuralism, idealization and approximation. Rodopi, Amsterdam, pp 127–139

Ory RL (ed.) (1981) Antinutrients and natural toxicants in food. Food & Nutrition Press, Wesport

Pacher M, Puchta H (2017) From classical mutagenesis to nuclease-based breeding – directing natural DNA repair for a natural end-product. Plant J 90: 819–833

Perry JN, ter Braak CJF, Dixon PM, Duan JJ, Hails RS, Huesken A, Lavielle M, Marvier M, Scardi M, Schmidt K, Tothmeresz B, Schaarschmidt F, van der Voet H (2009) Statistical aspects of environmental risk assessment of GM plants for effects on non-target organisms. Environ Biosafety Res 8: 65–78

Phelan SE (2001) What is complexity science really? Emergence 3: 120–136

Popper KR (2002) Conjectures and refutations. Fifth edition. Routledge, London

Popper KR (1979) Objective knowledge: an evolutionary approach. Second edition. Clarendon Press, Oxford

Prado JR, Segers G, Voelker T, Carson D, Dobert R, Phillips J, Cook K, Cornejo C, Monken J, Grapes L, Reynolds T, Martino-Catt S (2014) Genetically engineered crops from idea to product. Annu Rev Plant Biol 65: 769–790

Rayan AM, Abbott LC (2015) Compositional analysis of genetically modified corn events (NK603, MON88017 x MON810 and MON89034 x MON88017) compared to conventional corn. Food Chem 176: 99–105

Raybould A (2010) Reducing uncertainty in regulatory decision-making for transgenic crops: more ecological research or clearer environmental risk assessment? GM Crops 1: 25–31

Raybould A, Macdonald P (2018) Policy-led comparative environmental risk assessment of genetically modified crops: testing for increased risk rather than profiling phenotypes leads to predictable and transparent decision-making. Front Bioeng Biotechnol 6: 43, 1–7

Ricroch A, Bergé JB, Kuntz M (2011) Evaluation of genetically engineered crops using transcriptomic, proteomic and metabolic profiling techniques. Plant Physiol 155: 1752–1761

Ridley WP, Shillito RD, Coats I, Steiner H-Y, Shawgo M, Phillip A, Dussold P, Kurtyka L (2004) Development of the International Life Science Institute Crop Composition Database. J Food Compos Anal 17: 423–438

Romeis J, Bartsch D, Bigler F, Candolfi MP, Gielkens MMC, Hartley SE, Hellmich RL, Huesing JE, Jepson PC, Layton R, Quemada H, Raybould A, Rose RI, Schiemann J, Sears MK, Shelton AM, Sweet J, Vaituzis Z, Wolt, JD (2008) Assessment of risk of insect-resistant transgenic crops to nontarget arthropods. Nat Biotech 26: 203-208.

Sang JP, Salisbury PA (1988) Glucosinolate profiles of international rapeseed lines (*Brassica napus* and *Brassica campestris*). J Sci Food Agric 45: 255–261

Sanvido O, Romeis J, Gathmann A, Gielkens M, Raybould A, Bigler F (2012) Evaluating environmental risks of genetically modified crops – ecological harm criteria for regulatory decision-making. Environ Sci Pol 15: 82–91

Schaart JG, van der Wiel CCM, Lotz LAP, Smulders MJM (2016) Opportunities for products of new plant breeding techniques. Trends Plant Sci 21: 438–449

Stephens PA, Buskirk SW, del Rio, CM (2007) Inference in ecology and evolution. Trends Ecol Evol 22: 192–197

Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennett M (2017) Plant phenomics, from sensors to knowledge. Curr Biol 27: R770–R783

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science 327: 818–822

Thornton S (2018) Karl Popper. In: Edward N (ed) The Stanford encyclopedia of philosophy. https://plato.stanford.edu/entries/popper/

van der Voet H, Perry JN, Amzal B, Paoletti C (2011) A statistical assessment of differences and equivalences between genetically modified and reference plant varieties. BMC Biotechnol 11:15, 1–20

Vitashev S, Schwartz C, Lenderts B, Young JK, Cigan AM (2016) Genome editing in maize directed by CRISP-Cas9 ribonucleoprotein complexes. Nat Commun 7: 13274

Vlaanderen J, Moore LE, Smith MT, Lan Q, Zhang L, Skibola CF, Rothman N, Vermeulen R (2010) Application of omics technologies in occupational and environmental health research; current status and projections. Occup Environ Med 67: 136–143

US EPA (United States Environmental Protection Agency) (2007) Biopesticides registration action document. Modified Cry3A protein and the genetic material necessary for its production (via elements of pZM26) in Event MIR604 corn SYN-IR604-8. https://www3.epa.gov/pesticides/chem_search/reg_actions/registration/decision_PC-006509_1-Mar-07.pdf

Wolt JD (2017) Safety, security, and policy considerations for plant genome editing. Prog Mol Bio Transl Sci 149: 215–241

Wunderlich S, Gatto KA (2015) Consumer perception of genetically modified organisms and sources of information. Adv Nutr 6: 842–851

Wynne B (2001) Creating public alienation: expert cultures of risk and ethics on GMOs. Sci Cult (Lond) 10: 445–481

Xu Y, Crouch JH (2008) Marker assisted selection in plant breeding: from publications to practice. Crop Sci 48: 391–407

Yin K, Gao C, Qiu J-L (2017) Progress and prospects in plant genome editing. Nat Plants 3: 17107, 1–6

Zamir D (2001) Improving plant breeding with exotic gene libraries. Nat Rev Genet 2: 983–989

Zeller SL, Kalinina O, Brunner S, Keller B, Schmidt B (2010) Transgene x environment interactions in genetically modified wheat. PLoS One 5: e1140